



The Discovery, Characterization and Crystallographically Determined Binding Mode of an FMOC-Containing Inhibitor of HIV-1 Protease

Earl E. Rutenber,^a James J. De Voss,^b Lucas Hoffman,^a Robert M. Stroud,^{a,b} Kwan H. Lee,^b Juan Alvarez,^b Fiona McPhee,^b Charles Craik^{a,b} and Paul R. Ortiz de Montellano^{b,*}

^aDepartment of Biochemistry and Biophysics and ^bDepartment of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, CA 94143, U.S.A.

Abstract—A pharmacophore derived from the structure of the dithiolane derivative of haloperidol bound in the active site of the HIV-1 protease (HIV-1 PR) has been used to search a three-dimensional database for new inhibitory frameworks. This search identified an FMOC-protected *N*-tosyl arginine as a lead candidate. A derivative in which the arginine carboxyl has been converted to an amide has been crystallized with HIV-1 PR and the structure has been determined to a resolution of 2.5 Å with a final R-factor of 18.5%. The inhibitor binds in an extended conformation that results in occupancy of the S2, S1', and S3' subsites of the active site. Initial structure–activity studies indicate that: (1) the FMOC fluorenyl moiety interacts closely with active site residues and is important for binding; (2) the *N*^G-tosyl group is necessary to suppress protonation of the arginine guanidinyll terminus; and (3) the arginine carboxamide function is involved in interactions with the water coordinated to the catalytic aspartyl groups. FMOC-protected arginine derivatives, which appear to be relatively specific and nontoxic, offer promise for the development of useful HIV-1 protease inhibitors. © 1997 Elsevier Science Ltd.

Introduction

The pandemic spread of the human immunodeficiency virus (HIV-1), the causative agent of AIDS,¹ has led to an intensive scrutiny of its lifecycle for targets that are susceptible to chemotherapeutic intervention. One such target is the virally encoded protease (HIV-1 PR).² This homodimeric aspartyl protease is responsible for processing the polyprotein produced from transcription of the *gag* and *pol* genes into the individual proteins required for replication and assembly of the mature virus.³ Mutational inactivation of the protease by alteration of its catalytic aspartate residues has been shown to result in the production of noninfectious virions.⁴ The key role of the protease and the availability of high-resolution crystallographic structures of HIV-1 PR have made it a prime target for the rational design of anti-AIDS therapeutic agents.

A large number of potent inhibitors of the protease have been produced from the intensive search for anti-AIDS compounds, many of which have nanomolar and even subnanomolar *K_i* values.^{5,6} Many of these inhibitors are peptide derivatives of low clinical utility due to the low bioavailability and rapid degradation characteristic of peptidic therapeutics, but several

modified peptidic agents have proven successful in the clinic and are currently commercially available. Nevertheless, the development of resistance to the agents in clinical use and their high cost makes the search for simpler agents with improved pharmacokinetic properties and different binding interactions highly desirable.

We have been involved in the use of computational approaches to the identification of nonpeptidic structures as potential leads for the development of HIV-1 PR inhibitors. These studies have led to the identification of haloperidol,^{7–9} curcumin,¹⁰ and biliverdin/bilirubin¹¹ as lead structures and to the synthesis of derivatives of these compounds that inhibit HIV-1 PR. The development of effective inhibitors of HIV-1 PR based on the haloperidol skeleton (**2** in Fig. 1 is the dithiolane derivative of haloperidol) has been unsuccessful due, in part, to the toxicity of the parent compound. To extrapolate the knowledge gained from the binding of haloperidol derivatives in the protease active site to the identification of new inhibitor scaffolds, we have constructed a search pharmacophore. Here we report the discovery, characterization, and initial structure–activity analysis of a novel class of inhibitors of HIV-1 PR based on the protected arginine derivative *N*_α-FMOC-*N*_ω-tosylarginine (**1**, Fig. 1). We describe the X-ray crystallographic structure of the complex formed between HIV-1 PR and a derivative of **1** which reveals an extended binding mode that results in occupancy of the S2, S1', and S3' subsites of the enzyme active site.

*Corresponding author: Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, CA 94143-0446, U.S.A. Tel.: (415) 476-2903 fax: (415) 476-0688 e-mail: ortiz@cgl.ucsf.edu

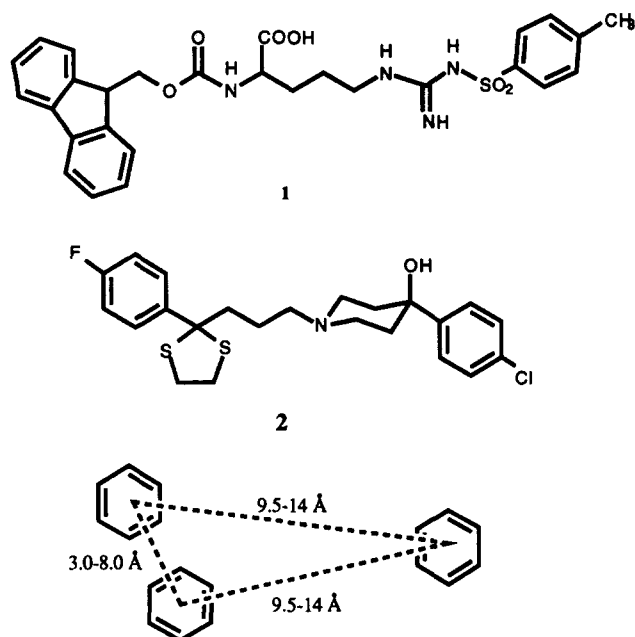


Figure 1. Structures of the haloperidol dithiolane derivative (**2**) used to develop the pharmacophore model and the lead Fmoc compound (**1**) identified by the search. The query used for the pharmacophore search of the Fine Chemicals Database with the program MACCS-3D is also shown.

Results and Discussion

Discovery

A pharmacophore based upon the crystallographically determined binding mode of the haloperidol dithiolane derivative (**2**)⁹ and the specific protein–ligand interactions thus revealed was defined (Fig. 1). Hydrophobic binding of the aromatic and dithiolane rings of **2** to the protein are of prime importance in determining the potency of the inhibitor, and their spatial arrangement formed the core of the pharmacophore, a triangle, with the centroid of an aromatic ring at each of the vertices. The centroid to centroid distances range from 9.5 to 14.0 Å and from 3.0 to 8.0 Å to allow for conformationally flexible molecules capable of satisfying the stipulated requirements. A search was performed of the available chemicals database (ACD) (Molecular Design), a collection of approximately 60,000 commercially available compounds, employing the program MACCS-3D (Molecular Design).¹² One-hundred-and-thirty-one molecules satisfied the query, including 30 Fmoc-protected aromatic amino acids, some of which also bear a second aromatic protecting group. This led to the testing of a dozen Fmoc-protected amino acids, the most promising of which, in terms of IC_{50} values, were derivatives of arginine (**1**, IC_{50} 5 μ M).

The Fmoc arginine compound **1** also inhibits cathepsin D but is quite specific as an inhibitor for HIV-1 PR in comparison to the mammalian aspartyl proteases pepsin and renin (Table 1). The toxicities of **1** and **7** (Table 2) are less than that of the haloperidol thioketal **2**. Furthermore, unlike **2**, **1** shows no salt dependence

Table 1. Specificity of **1** as an inhibitor of different aspartyl proteases

Enzyme	IC_{50} (μ M)
HIV-1 PR	5
Pepsin	>500
Renin	$\gg$ 500
Cathepsin D	12

Table 2. IC_{50} and LD_{50} values for **1** and **8** and for the haloperidol derivative **2**

Inhibitor	IC_{50} (μ M)	LD_{50} (μ M)
1	5	>500
7	8	50
8	4	275
9	12	40
12	9	100
2	15	20

on inhibition and effectively inhibits HIV-1 PR at physiological salt concentrations.

Three-dimensional binding mode

To define the binding mode of these inhibitors and to develop specific strategies for further inhibitor design and improvement, HIV-1 PR was co-crystallized with inhibitors in this series. Compound **8**, the amide derivative of **1**, yielded large diffraction quality crystals and its three-dimensional crystallographic structure bound to HIV-1 PR was determined to a resolution of 2.5 Å (Figs 2 and 3).

The inhibitor lies within the active site in a relatively extended conformation (Fig. 2) filling some of the volume occupied by substrate-based peptidomimetic inhibitors (Fig. 4). The difference in electron density for **8** shows clear and detailed features which allow unambiguous placement and successful refinement of the inhibitor model. Refinement of the inhibitor model in the twofold related orientation generated using the C2 symmetry of the protein resulted in a poorer fit between the model and the electron density. This was principally in the region of the large aromatic fluorenyl ring, most of which had no electron density in the twofold related orientation. The individual B-factors for the atoms in the structure are refined. The average B-factor for nonhydrogen protein atoms is 21.3 Å² and the average B-factor for nonhydrogen inhibitor atoms is 38.4 Å².¹³

The protein C α atoms were used to guide the least squares superposition of the **8**-bound HIV-1 PR structure and the 7HPV structure¹⁴ from the Protein Data Bank¹⁵ to compare the binding mode of **8** to that of the peptidomimetic inhibitor JG365. The C α atoms of the two structures superimpose with an RMS deviation of 0.56 Å and all nonhydrogen atoms of the two structures superimpose with an RMS deviation of

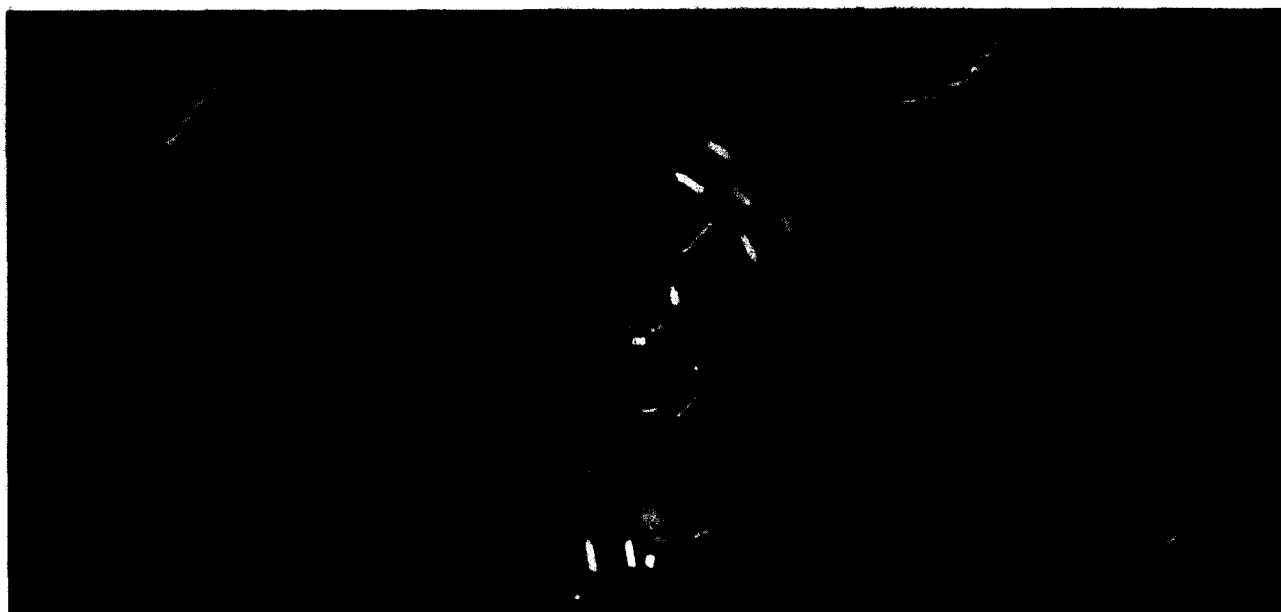


Figure 2. The inhibitor **8** (colored by atom type) bound to the active site of HIV-1 PR with one monomer colored cyan and the other colored magenta. The view is along the dimer twofold into the active site with the flap residues, 45–55 and 145–155, removed for clarity. The symmetrically disposed catalytic aspartic acid residues, 25 and 125, lie below the bound inhibitor and are colored red. Hydrogen bonds (colored yellow) and liganded water molecules are shown in detail in Figures 5 and 6.

1.14 Å. Compound **8** and the peptidomimetic inhibitor JG365 from the 7HVP structure are shown after superposition in Figure 4. The overlapped structures reveal that **8** binds within the active site to fill subsites S2, S1' and S3' with excursions toward subsites S3 and

S4' leaving subsites S1 and S2' vacant. The amide nitrogen adjacent to the C α of the arginyl residue coordinates a water molecule bound between the active site aspartic acid residues. The carboxamide oxygen at the modified carboxy terminus of the arginyl residue

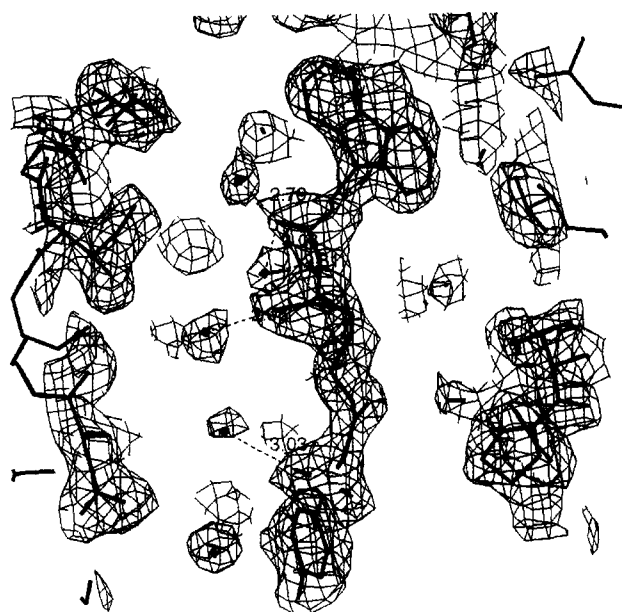


Figure 3. The bound conformation of **8** and well-ordered water molecules within the active site. The atoms are colored by type: oxygen is red, nitrogen is blue, carbon is green, and hydrogen is white. The electron density map for the inhibitor, bound water molecules, and surrounding protein is shown at a 1.0 sigma contour level. The view is approximately that shown in Figure 2.

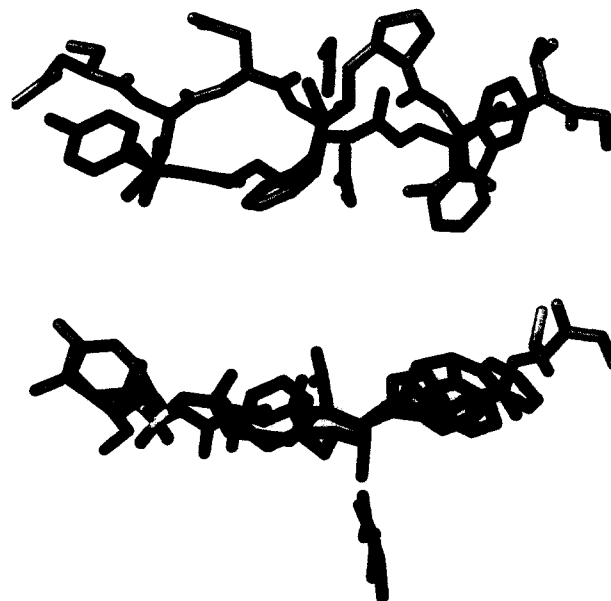


Figure 4. Two views of **8** and JG365 after superimposition guided by the protein C α s of the two structures. The view on the left is along the enzyme dimer twofold (similar to the view in Fig. 2) with the catalytic aspartic acid residues included for orientation, and shows the disposition of the inhibitor **8** relative to the functional groups of JG365 which occupy the S3–S3' subsites in the HIV-1 PR active site. The view on the right is perpendicular to the dimer twofold.

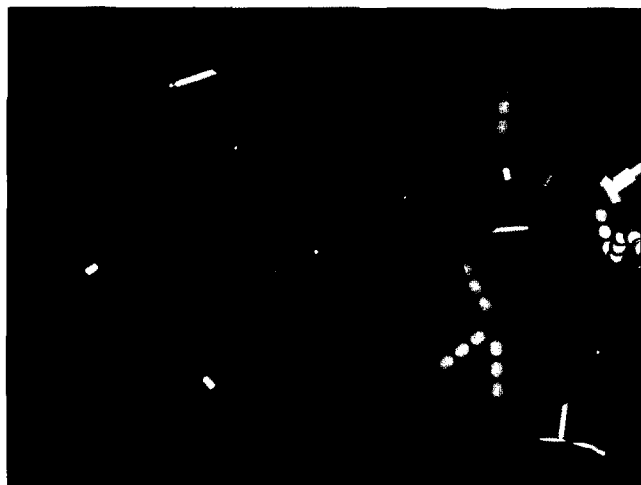


Figure 5. Close-up view of the fluorenyl ring of **8** showing some of the residues that comprise the hydrophobic binding pocket.

displaces the water molecule bound between the flap residues in unliganded structures.

The fluorenyl ring is the most well-ordered portion of the inhibitor with atomic B factors ranging from 29.6 to 32.6 Å². The fluorenyl ring is bound mostly within the S2 pocket of the enzyme active site, with one edge of the ring system extending into the S3 binding pocket (Fig. 5). This binding site is formed by the backbone and hydrophobic side-chain atoms from a loop of residues (A128, D129, D130, T131, and V132) and by the side-chain atoms of residues I147 and I184. Extensive van der Waals interactions between nonpolar protein atoms secures the aromatic fluorenyl ring within the S2 and S3 subsites.

The arginyl moiety of **8** extends from the S2 binding site over the catalytic aspartate residues. A water molecule, Wat 379, is coordinated by the protected amino nitrogen of the arginine residue and the OD1 carboxylate oxygens of residues D25 and D125 (Figs 5 and 6). The aliphatic side-chain of the bound arginyl portion of **8** extends into the S1' binding site, leaving the S1 pocket unoccupied. The carboxamide carbonyl of **8** is hydrogen bonded to one of the amide nitrogens of I50 which, together with the amide nitrogen of I150, normally coordinates a water molecule (Figs 5 and 6). This water molecule is displaced by the carboxamide carbonyl. The carboxamide NH₂ group extends toward the S2' subsite and ligates a water molecule which makes no hydrogen bonds to the protein (Fig. 3).

The tosyl group of **8** is bound in the S3' site, leaving the S2' site vacant. The sulfonamide oxygens form a bifurcated hydrogen bond to the terminal NH₂ of R108 (Fig. 3). Other atoms of the tosyl moiety make no specific interaction with the protein, and most are solvent exposed. The atomic B-factors for the tosyl group of **8** range from 43.0 to 46.1 Å² indicating that this portion of the inhibitor is the least well ordered.

A comparison of the binding modes determined for **2'** and **8** bound to the active site of HIV-1 PR reveals that

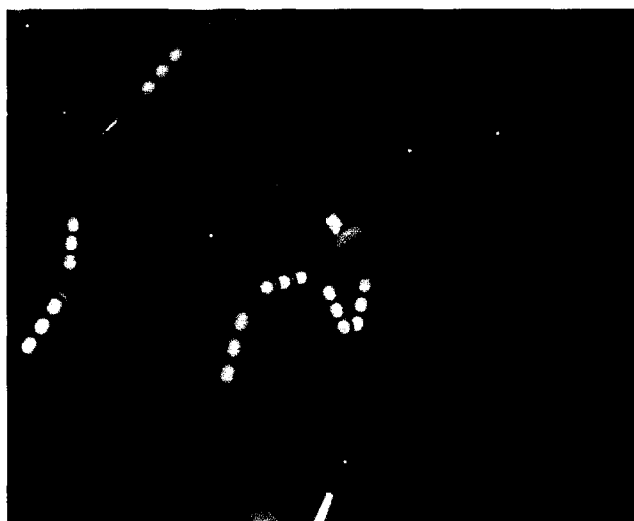


Figure 6. Close-up view of the tosylate-protecting moiety and the hydrogen bonds between the sulfonate oxygens and the terminal NH₂ of R108.

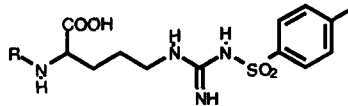
the aromatic ring systems targeted in the pharmacophore search do not overlap in the expected manner. Both compounds fill similar volumes in the active site, but **8** is more extended and lies at the outside of the allowed distance between the aromatic rings. This discrepancy between the binding modes of the two inhibitors is unexpected but not surprising in view of the flexibility of the compounds and the fact that **2** itself has been found to bind in two different orientations.⁹

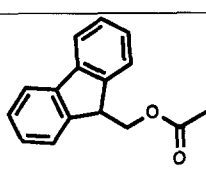
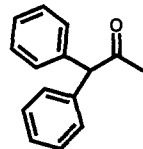
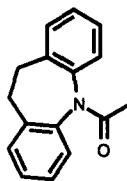
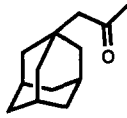
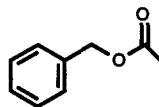
Structure-activity relationships

To clarify the binding interactions of **1** with HIV-1 PR, analogues of **1** were examined with modifications of: (1) the α-amino-protecting group (**3–6**); (2) the α-carboxyl group (**7–13**); (3) the guanidino moiety (**14** and **15**); and (4) the sulfonyl-protecting group (**16–21**). Examination of the IC₅₀ values of these compounds leads to the following general conclusions.

The fluorenyl ring system is important for specific binding since alterations at this site lead to at least an order of magnitude increase in IC₅₀ (Table 3). This is consistent with the crystallographic finding that the fluorenyl moiety is the most well-ordered part of the structure.

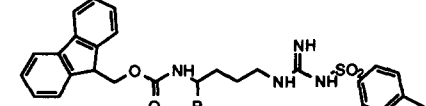
Modifications at the α-carboxylate site have mixed effects. Substitutions of small functional groups are well tolerated (compounds **1**, **7**, **8**, **9** and **12**) whereas larger more extended functional groups (compounds **10** and **11**) lead to an increase in IC₅₀ (Table 4). It is clear that this group is involved in energetically important interactions with the protein as changing the absolute configuration of the arginine from *S* to *R* results in a large increase in IC₅₀ (compound **13**), perhaps due to abrogation of the key interaction between the carboxamide carbonyl oxygen and the flap backbone amide nitrogen.

Table 3. Structures and HIV-1 PR inhibitory activities of arginine derivatives modified at the α -amino group


Compound	R	IC ₅₀
1		5
3		45
4		220
5		54
6		>>500

An important function of the sulfonyl-protecting moiety seems to be lowering the pK_a of the guanidine group so that it is no longer protonated. Thus, the nonprotonated N_α -FMOC-citrulline derivative **14** (IC₅₀ = 30 μ M) binds much better than the isosteric but charged N_α -FMOC-arginine **15** (IC₅₀ = 580 μ M) and binds almost as well as the uncharged sulfonyl arginine **1** (Table 5). The structure of the aromatic moiety of the sulfonyl group does improve the binding, but no structure-activity relationship is apparent (Table 6), perhaps because the substituent on the sulfonyl group exhibits no specific interactions with the protein. However, the toluene-sulfonyl moiety appears optimal among those investigated.

Analysis of the interaction of the inhibitor with the protein provides a rationalization of the observed IC₅₀ values (Fig. 7). The fluorenyl ring of the FMOC moiety binds in the S2 and S3 pockets of the active site, interacting with hydrophobic residues, including V132, I147, and I184. This is clearly an important binding

Table 4. Structures and HIV-1 PR inhibitory activities of N_α -FMOC- N_ω -tosylarginine derivatives modified on the carboxyl group


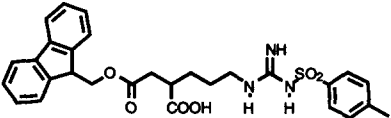
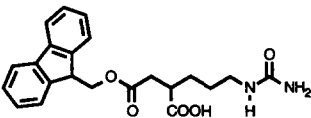
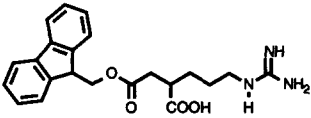
Compound	R	IC ₅₀ (μ M)
1	CO ₂ H	5
7	CO ₂ CH ₃	8
8	CONH ₂	4
9	CO ₂ CH ₂ CH ₃	12
10	CO ₂ CH ₂ C ₆ H ₅	60
11	CO ₂ CH ₂ CH ₂ OH	55
12	CH ₂ OH	9
13	CO ₂ H (R isomer)	>100

interaction as changing the fluorenyl ring to a phenyl group dramatically raises the IC₅₀ (Table 3; **6** 500 μ M). A range of other modifications of the FMOC moiety (**3**–**5**) also adversely affects the IC₅₀, indicating the complementarity of this group with the protease.

The carboxamide carbonyl of **8** is hydrogen bonded to the α -NH of I50 in the crystal structure of its complex with the protease. This hydrogen bond replaces the water that is normally seen between the two flaps in the closed form of the protease, where it forms a hydrogen-bonded complex with the NHs of I50 and I150. The carboxamide NH₂ of **8** appears oriented toward the substrate channel of the active site and does not interact with the protein. This binding mode is reflected in the IC₅₀s of the various analogues (**7** and **9**–**12**). Changing the substituent at the α -carboxyl to a small functional group appears to have little effect on the IC₅₀ as these derivatives maintain the H-bond acceptor for interaction with the α -NH of I50. Only bulky or extended substituents appear to interact adversely with the protease. This interaction is clearly an important one, however, as changing the configuration of the α -carbon from *S* to *R* (**13**) raises the IC₅₀ by over an order of magnitude. The one anomalous result in this series is **11**, where the hydroxyethyl moiety significantly decreases the affinity of the compound for the protease. One potential explanation for this is that the hydroxyl group of the ester side chain forms an intramolecular hydrogen bond with the carbonyl in competition with the α -NH of I50, thus lowering the affinity of the inhibitor for the protease.

The toluenesulfonyl guanidinyll moiety of **8** appears in the crystal structure to interact only with the protease through the sulfonyl group itself. The aromatic ring appears to be located mainly in bulk solvent at the periphery of the active site and the guanidine fragment of **8** does not interact directly with the protein. The major interaction between this part of the inhibitor and the protease is through hydrogen bonds from the oxygens of the sulfonyl group to a terminal NH₂ of

Table 5. Structures and HIV-1 PR inhibitory activities of arginine derivatives modified at the imino group

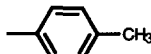
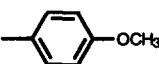
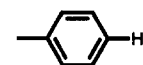
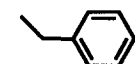
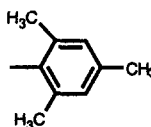
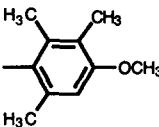
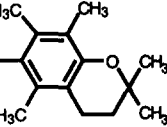
Compound	IC ₅₀
 5	
 30	
 580	

R108. Charge repulsion between R108 and the protonated guanidine of **15** explains the large difference in IC₅₀ between **14** and the uncharged analogue **15**. The structure supports the notion that a major role of the sulfonyl-protecting group is to lower the pK_a of the guanidine, preventing this unfavorable charge-charge interaction. The lack of correlation between the structure of the arylsulfonyl moiety and the observed IC₅₀ agrees with the lack of specific interaction of this group and the protein in the crystal structure (Table 6). Nonspecific interactions presumably occur in solution which lead to the observed differences in IC₅₀.

Conclusions

We have presented a novel series of HIV-1 PR inhibitors based upon an FMOC/tosylate-protected arginine residue. Our initial lead compound **1** displays micromolar affinity for the enzyme as well as low toxicity and good specificity. The binding interactions of the carboxamide derivative **8** with the protease have been delineated by a crystallographic structure of **8** complexed with the protease. Preliminary structure-activity relationships agree with those expected from the crystal structure. The fluorenyl moiety interacts closely with hydrophobic residues in the S2 and S3 pockets; replacing it decreases the inhibitor's affinity for the enzyme. Hydrogen bonds are formed between the carboxamide carbonyl oxygen of **8** and a protein backbone NH (I50), the amide nitrogen of **8** and the water coordinated between the catalytic aspartic acids, and the sulfonyl oxygens of **8** and the terminal NH₂ of R108. The hydrogen bond to I50 would appear to be particularly important as the water molecule normally found coordinated to it and I50 has been displaced. A major function of the protecting sulfonyl group appears to be suppression of the charge on the guanidinium

Table 6. Structures and HIV-1 PR inhibitory activities of arginine derivatives modified at the N_ω-amino group

Compound	R	IC ₅₀ (μM)
1		5
16		40
17		25
18		370
19		70
20		66
21		30

moiety. This scaffold can be used to design further reversible and irreversible inhibitors of the HIV-1 PR.¹⁷

A comparison of the crystallographic binding modes of **8** (the carboxamide derivative of **1**) and **2** reveals that while the fluorenyl ring of **8** and the chlorophenyl moiety of **2** bind in the same hydrophobic pocket, there is little else that is the same in their interactions with the protease.

Experimental

General procedure

Compounds **1**, **6**, **13**, **14**, **15**, **18**, **19**, **20** and **21**, and the protected arginines used for the synthesis of other analogues, were purchased from Bachem Bioscience

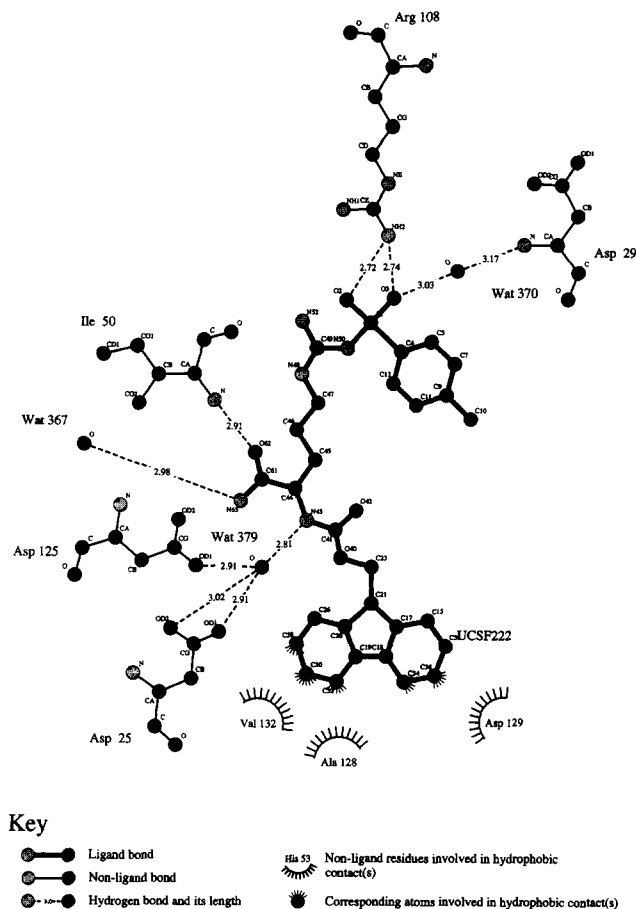


Figure 7. Schematic diagram of the ligand-binding determinants as displayed by the program LIGPLOT.¹⁶

(King of Prussia, PA). Melting points were determined by the capillary method and are uncorrected. Flash column chromatography was performed on silica gel (230–400 mesh). Proton and carbon NMR spectra were recorded on a General Electric QE 300 instrument. Mass spectra were obtained on a VG-70 spectrometer.

***N*^α-Fmoc-*N*^ε-tosylarginine methyl ester (7).** The methyl ester **7** was formed quantitatively by the reaction of **1** with diazomethane. Thus, the acid **1** (18 mg, 0.033 mmol) was dissolved in methanol and an excess of ethereal diazomethane was added. After stirring at 25 °C for 30 min the solvent was evaporated to give **7** as a foam (19 mg) that was purified by column chromatography (SiO₂, 70% ethyl acetate:hexane): ¹H NMR (300 MHz, CDCl₃) δ 1.54–1.78 (brd m, 4 H, CH₂CH₂CH₂N), 2.32 (s, 3H, CH₃C₆H₄), 3.16–3.25 (brd m, 2H, CH₂CH₂CH₂N), 3.69 (s, 3H, CH₃O), 4.15 (t, 1H, *J* = 6.9 Hz, CHCH₂O), 4.26–4.30 (m, 1H, CHCO₂), 4.35 (d, 2H, *J* = 6.9 Hz, CHCH₂O), 5.72 (brd d, 1H, *J* = 4.5 Hz, CHNH), 6.34 (brd s, 2 H, 2 × NH), 7.14–7.78 (m, 12 H, aromatic Hs); ¹³C NMR (75 MHz, CDCl₃) δ 21.36, 25.06, 29.99, 40.66, 47.06, 52.53, 53.26, 67.10, 119.96, 125.02, 125.89, 127.07, 127.72, 129.20, 140.60, 141.23, 142.04, 143.56, 143.67, 156.41, 156.78, 172.67; MS (LSIMS), *m/z* 587 (M+Na⁺), 565 (MH⁺).

***N*^α-Fmoc-*N*^ε-tosylargininecarboxamide (8).** The acid **1** (61 mg, 0.11 mmol) was dissolved in dry CH₂Cl₂ (3 mL) and oxalyl chloride (15 μL, 1.5 equiv) was added along with a catalytic amount of DMF. After stirring at 0 °C for 30 min an excess of methanol saturated with ammonia (3 mL) was added and the reaction was stirred for a further 2–3 min before being quenched with 1 N HCl (10 mL). It was then diluted with ethyl acetate (30 mL) and the organic layer was separated and successively washed with satd NaHCO₃ solution (10 mL) and brine (10 mL), dried and evaporated under reduced pressure. The residue was purified by flash column chromatography (SiO₂, 4:1 acetone:hexane) to give the amide **8** (40 mg, 65.7%): ¹H NMR (300 MHz, CDCl₃) δ 1.53–1.80 (brd m, 4H, CH₂CH₂CH₂N), 2.23 (s, 3H, CH₃C₆H₄), 3.12–3.25 (brd m, 2H, CH₂CH₂CH₂N), 4.05–4.24 (m, 4H, CHCH₂O and CHCO₂), 6.53–6.64 (brd m, 5H, 5 × NH), 7.06–7.69 (m, 12 H, aromatic Hs); ¹³C NMR (75 MHz, CDCl₃) δ 21.31, 25.57, 29.35, 41.0, 46.98, 54.34, 67.02, 119.85, 125.14, 125.86, 127.05, 127.64, 129.28, 140.20, 141.12, 142.23, 143.63, 143.71, 156.64, 157.00, 175.67; MS (LSIMS), *m/z* 572 (M+Na⁺), 550 (MH⁺).

General procedure for the synthesis of *N*^α-Fmoc-*N*^ε-tosylarginine esters 9–11

The acid **1** (55 mg, 0.1 mmol) was dissolved in dry CH₂Cl₂ (3 mL) and oxalyl chloride (10 μL, 1.1 equiv) was added along with a catalytic amount of DMF. After stirring at 25 °C for 30 min an excess of the appropriate alcohol (as specified) and pyridine (20 μL, 2 equiv) were added and the reaction was stirred for 2 h. The mixture was then diluted with ethyl acetate (30 mL) and washed successively with 1 N HCl (10 mL), satd NaHCO₃ solution (10 mL), and brine (10 mL). The resulting solution was dried over anhydrous MgSO₄ and evaporated under reduced pressure. The residue was purified by flash column silica gel chromatography (solvent as specified below) to give the esters **9–11** in the indicated yields:

***N*^α-Fmoc-*N*^ε-tosylarginine ethyl ester (9).** Ethanol–70% ethyl acetate/hexane (v/v), 54% yield: ¹H NMR (300 MHz, CDCl₃) δ 1.22 (t, 3H, *J* = 7.5 Hz, CH₃CH₂O), 1.54–1.85 (brd m, 4H, CH₂CH₂CH₂N), 2.32 (s, 3H, CH₃C₆H₄), 3.10–3.30 (brd m, 2H, CH₂CH₂CH₂N), 4.11–4.18 (m, 3H, CHCH₂O and OCH₂CH₃), 4.24–4.30 (m, 1H, CHCO₂), 4.34 (d, 2H, *J* = 6.6 Hz, CHCH₂O), 5.74 (brd s, 1H, CHNH), 6.37 (brd s, 2H, 2 × NH), 7.14–7.75 (m, 12H, aromatic Hs); ¹³C NMR (75 MHz, CDCl₃) δ 14.13, 21.36, 25.04, 30.13, 40.70, 47.06, 53.27, 61.72, 67.11, 119.96, 125.03, 125.91, 127.07, 127.72, 129.20, 140.54, 141.22, 142.05, 143.56, 143.67, 156.44, 156.75, 172.21; MS (LSIMS), *m/z* 601 (M+Na⁺), 579 (MH⁺).

***N*^α-Fmoc-*N*^ε-tosylarginine benzyl ester (10).** Benzyl alcohol–70% ethyl acetate/hexane (v/v), 38% yield: ¹H NMR (300 MHz, CDCl₃) δ 1.57–1.91 (brd m, 4H, CH₂CH₂CH₂N), 2.38 (s, 3H, CH₃C₆H₄), 3.10–3.30 (brd

m, 2H, CH₂CH₂CH₂N), 4.22 (t, 1H, *J* = 7.2 Hz, CHCH₂O), 4.35–4.46 (m, 3H, CHCO₂ and CHCH₂O), 5.20 (s, 2H, C₆H₅CH₂), 6.84 (brd s, 1H, CHNH), 6.40 (brd s, 2H, 2 × NH), 7.19–7.83 (m, 17H, aromatic Hs); ¹³C NMR (75 MHz, CDCl₃) δ 21.33, 25.04, 29.94, 40.64, 47.07, 53.37, 67.12, 67.33, 119.94, 125.02, 125.88, 127.05, 127.71, 128.24, 128.52, 128.62, 129.19, 135.05, 140.64, 141.23, 142.01, 143.57, 143.67, 156.39, 156.76, 172.02; MS (LSIMS), *m/z* 663 (M+Na⁺), 641 (MH⁺).

N^α-FMOC-N^G-tosylarginine hydroxyethyl ester (11). Ethylene glycol–ethyl acetate, 28% yield: ¹H NMR (300 MHz, CDCl₃) δ 1.56–1.94 (brd m, 4H, CH₂CH₂CH₂N), 2.32 (s, 3H, CH₃C₆H₄), 3.10–3.25 (brd m, 2H, CH₂CH₂CH₂N), 3.75 (brd s, 2H, CH₂CH₂OH), 4.11–4.35 (m, 6H, CHCO₂CH₂CH₂ and CHCH₂O), 6.03 (d, 1H, *J* = 7.6 Hz, CHNH), 6.39 (brd s, 2H, 2 × NH), 7.15–7.74 (m, 12H, aromatic Hs); ¹³C NMR (75 MHz, CDCl₃) δ 21.39, 24.97, 29.37, 40.66, 47.11, 53.79, 60.40, 66.85, 67.11, 119.97, 125.08, 125.91, 127.10, 127.74, 129.31, 140.38, 141.26, 142.29, 143.63, 143.77, 156.44, 156.90, 172.46; MS (LSIMS), *m/z* 617 (M+Na⁺), 595 (MH⁺).

N^α-FMOC-N^G-tosylargininol (12). The methyl ester **7** (60 mg, 0.11 mmol) was dissolved in ethanol (2 mL) at –10 °C and calcium chloride (65 mg, 0.44 mmol) in dry THF (1 mL) and sodium borohydride (16.6 mg, 0.44 mmol) were added. After stirring for 5 h, the reaction was poured into 1 M citric acid and the mixture was extracted with ethyl acetate (3 × 30 mL). The combined extracts were washed successively with satd sodium bicarbonate solution (10 mL) and brine (10 mL) and were then dried and evaporated under reduced pressure. The residue was purified by flash column chromatography (SiO₂, 5% acetone:hexane) to give the alcohol **12** (21 mg, 37%). ¹H NMR (300 MHz, CDCl₃) δ 1.30–1.50 (brd m, 4H, CH₂CH₂CH₂N), 2.29 (s, 3H, CH₃C₆H₄), 3.05–3.25 (brd m, 2H, CH₂CH₂CH₂N), 3.25–3.60 (brd m, 3H, CHCH₂OH), 4.09 (t, 1H, *J* = 6.5 Hz, CHCH₂O), 4.29 (d, 2H, *J* = 6.8 Hz, CHCH₂O), 5.64 (brd s, 1H, CHNH), 6.42 (brd s, 2H, 2 × NH), 7.11–7.71 (m, 12H, aromatic Hs); ¹³C NMR (75 MHz, CDCl₃) δ 21.35, 25.55, 29.66, 41.97, 47.13, 64.61, 66.66, 119.90, 125.07, 125.86, 127.05, 127.66, 129.26, 140.29, 141.19, 142.24, 143.73, 143.78, 156.86, 156.98.

General procedure for preparation of arylsulfonyl-L-arginine derivatives

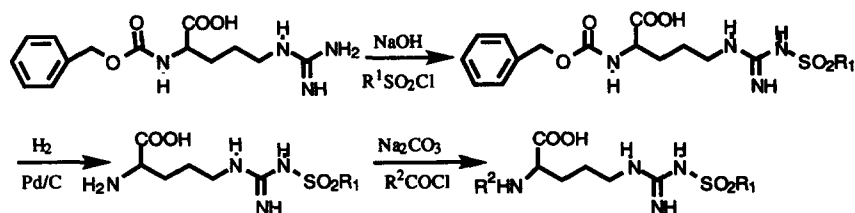
The procedure outlined in Scheme 1 was employed. A solution of arylsulfonyl chloride (60 mmol) in 80 mL of

acetone was added dropwise at 0 °C with vigorous stirring to a solution of carbobenzoxy-L-arginine (50 mmol) in a mixture of 50 mL of 4 M NaOH and 400 mL of acetone. After stirring for 2 h at 0 °C, the mixture was acidified with citric acid. The acetone was removed by evaporation in vacuo and the residue was extracted with ether (2 × 50 mL) to remove excess arylsulfonyl chloride. Ethyl acetate (200 mL) was added to the residue and the organic layer was washed with water (2 × 100 mL) and then extracted into 5% NaHCO₃ solution. After being acidified with citric acid, the oily solution was extracted with ethyl acetate (2 × 150 mL) and the combined extracts were washed with water (3 × 100 mL). The organic layer was dried over anhydrous MgSO₄ and then concentrated in vacuo. Flash column chromatography (9:1 CH₂Cl₂:CH₃OH) gave nearly colorless syrups in 50–70% yields.

The crude N^α-carbobenzoxy-N^G-arylsulfonyl-L-arginine in dry methanol (100 mL) was hydrogenated over Pd catalyst (2 g) for 24 h. The Pd catalyst was filtered off and the filtrate was evaporated in vacuo to dryness. Flash column chromatography (9:1 CH₂Cl₂:CH₃OH) gave the arylsulfonyl-L-arginine derivatives as white solids in yields of 30–45%. Benzenesulfonyl-L-arginine: mp 85–87 °C; ¹H NMR (DMSO-*d*₆, 300 MHz) δ 7.07 (m, 3H, aromatics Hs), 6.80 (d, *J* = 5.4 Hz, 2H, aromatic Hs), 2.8 (m, 2H, CH₂CH₂NH), 2.65 (m, 1H, CHCOOH), 0.8–1.7 (m, 4H, CH₂CH₂CH₂NH); ¹³C NMR (DMSO-*d*₆, 300 MHz) δ 172.14, 157.17, 144.60, 131.02, 128.62, 125.55, 53.77, 48.61, 39.50, 39.22; MS (CI), *m/z* 315 (MH⁺) 297, 285, 271, 200, 175, 115, 94. Benzenesulfonyl-L-arginine: mp 90–94 °C; ¹H NMR (DMSO-*d*₆, 300 MHz) δ 6.78 (m, 4H, aromatic Hs), 4.1 (s, 2H, CH₂C₆H₅), 3.05 (m, 2H, CH₂CH₂NH), 2.69 (m, 1H, CHCOOH), 0.9–1.9 (m, 4H, CH₂CH₂CH₂NH); ¹³C NMR (DMSO-*d*₆, 300 MHz) δ 171.47, 157.16, 131.62, 130.74, 127.91, 127.53, 68.79, 58.98, 48.59, 40.05, 25.70.

General procedure for N^α-alkyloxycarbonyl-N^G-arylsulfonyl-L-arginine

The desired alkylhaloformate (as specified below; 1.1 equiv) in dioxane (1 mL) was added to a solution of the arylsulfonyl-L-arginine (100 mg) in water (1 mL) containing Na₂CO₃ (2 equiv) (Scheme 1). The mixture was then stirred for 2–8 h before the solvent was removed in vacuo. Flash column chromatography (9:1 CH₂Cl₂:CH₃OH) gave the desired N^α-alkyloxycarbonyl-N^G-arylsulfonyl-L-arginines as white solids in yields of 35–55%.



Scheme 1. Synthetic scheme used for the preparation of arginine derivatives.

***N*^α-(Diphenylmethylcarbonyl)-toluenesulfonyl-L-arginine (3).** Diphenylmethylchloroformate—mp 163–165 °C; ¹H NMR (DMSO-*d*₆, 300 MHz) δ 7.7 (d, *J* = 7.8 Hz, 2H, aromatic Hs), 7.45 (m, 12H, aromatic Hs), 5.3 (s, 1H, NH), 4.1 (s, 1H, CH(C₆H₅)₂), 3.15 (m, 2H, CH₂NH), 2.45 (m, 1H, CHCOOH), 2.4 (s, 3H, CH₃C₆H₅), 1.4–1.9 (m, 4H, CH₂CH₂CH₂NH); ¹³C NMR (DMSO-*d*₆, 300 MHz) δ 170.16, 156.73, 141.75, 140.94, 140.84, 129.03, 128.67, 128.46, 128.14, 126.45, 125.56, 56.37, 53.63, 29.87, 25.56, 20.88; MS (LSIMS), *m/z* 545 (M+Na⁺), 523 (MH⁺).

***N*^α-(10,11-Dihydro-5H-dibenz[b,f]azepine-5-carbonyl)-toluenesulfonyl-L-arginine (4).** 10,11-Dihydro-5H-dibenz[b,f]azepine-5-chloroformate—mp 205–209 °C; ¹H NMR (DMSO-*d*₆, 300 MHz) δ 7.65 (d, 2H), 7.15 (m, 10H), 5.6 (d, *J* = 6.5 Hz, 2H), 4.15 (d, *J* = 6.5 Hz, 2H), 3.0 (m, 2H), 2.3 (m, 1H), 2.25 (s, 3H), 1.3–1.8 (m, 4H); ¹³C NMR (DMSO-*d*₆, 300 MHz) δ 175.82, 156.74, 154.99, 142, 140.98, 136.98, 130.13, 129.05, 127.46, 126.75, 125.58, 54.29, 30.10, 25.05, 20.90; MS (LSIMS), *m/z* 572 (M+Na⁺), 550 (MH⁺).

***N*^α-(1-Adamantylloxycarbonyl)-toluenesulfonyl-L-arginine (5).** Adamantylloxycarbonylchloroformate—mp 215–220 °C; ¹H NMR (DMSO-*d*₆, 300 MHz) δ 7.8 (d, *J* = 7.8 Hz, 2H, aromatic Hs), 7.5 (d, *J* = 7.75 Hz, 2H, aromatic Hs), 6.0 (brs, 2H), 2NH), 5.3 (brs, 1H, =NH), 3.2 (m, 2H, CH₂NH), 2.65 (m, 1H, CHCOOH), 2.2 (s, 3H, CH₃C₆H₅), 1.3–2.1 (m, 19H); ¹³C NMR (DMSO-*d*₆, 300 MHz) δ 170.5, 158.5, 156.142, 129.2, 125.5, 79.5, 46.5, 41.38, 41.2, 41.1, 35.9, 35.7, 30.1, 20.9; MS (LSIMS), *m/z* 529 (M+Na⁺), 484, 461, 431, 405.

***N*^α-(9-Fluorenylmethyloxycarbonyl)-*p*-methoxybenzenesulfonyl-L-arginine (16).** Mp 171–173 °C; ¹H NMR (DMSO-*d*₆, 300 MHz) δ 7.88 (d, *J* = 7.35 Hz, 2H, aromatic Hs), 7.66 (m, 4H, aromatic Hs), 7.36 (m, 4H, aromatic Hs), 6.99 (d, *J* = 7.4 Hz, 2H, aromatic Hs), 4.2 (m, 2H, CH₂O), 3.7 (s, 3H, OCH₃), 3.0 (m, 2H, CH₂NH), 2.5 (m, 1H, CHCOOH), 1.3–1.75 (m, 5H); ¹³C NMR (DMSO-*d*₆, 300 MHz) δ 161.09, 156.67, 155.49, 143.94, 143.87, 140.69, 127.59, 127.51, 127.07, 125.23, 120.08, 113.70, 65.35, 55.40, 46.73, 30, 27.5; MS (LSIMS), *m/z* 589.3 (M+Na⁺), 567 (MH⁺), 537, 469, 447, 433.

***N*^α-(9-Fluorenylmethyloxycarbonyl)benzenesulfonyl-L-arginine (17).** Mp 168–170 °C; ¹H NMR (DMSO-*d*₆, 300 MHz) δ 7.3–7.88 (m, 13H, aromatic Hs), 6.85 (brs, 2H, 2NH), 4.2 (m, 2H, CH₂O), 3.0 (m, 2H, CH₂NH), 2.5 (m, 1H, CHCOOH), 1.3–1.8 (m, 4H, CH₂CH₂CH₂NH); ¹³C NMR (DMSO-*d*₆, 300 MHz) δ 178, 156.82, 155.55, 143.96, 143.87, 140.71, 131.05, 128.93, 127.60, 125.51, 121.38, 120.08, 109.76, 65.33, 55.5, 46.77, 30, 27.5; MS (LSIMS), *m/z* 559 (M+Na⁺), 537 (MH⁺), 513, 482, 466.

In vitro assays of HIV-1 PR inhibitors

HIV-1 PR was assayed against the decapeptide Ala-Thr-Leu-Asn-Phe-↓Pro-Ile-Ser-Pro-Trp, which corre-

sponds to the HIV-1 carboxy-terminal autoprocessing site.¹⁸ The decapeptide was synthesized by conventional solid-state methods. Reactions were carried out as previously described.⁷ Conversion of the decapeptide to the two pentapeptides was quantified by HPLC and compared with product standard curves. The IC₅₀ determinations were carried out at pH 5.5. Stock solutions of the inhibitors (5 μM–5 mM) were prepared in DMSO. Compounds were added to assay solutions and the final DMSO concentration adjusted to 5% (v/v). Baseline values were determined from enzymatic reactions containing 5% DMSO in the absence of inhibitor. The enzyme concentration ranged typically from 1 to 4 μg/mL per assay. HIV-1 PR was pre-incubated with the respective inhibitor for 1 min at 25 °C before initiating the reaction by adding the decapeptide substrate (250 μM final concentration). The extent of inhibition is independent of the incubation time, indicating reversible binding. The activities of the inhibitors with respect to pepsin, renin, and cathepsin D were assayed as previously described.⁷ Toxicity assays on the inhibitor were carried out as described previously using a stable cell line (CH-1) that produces all of the HIV-1 HXB2 proteins with the exception of envelope gp160.¹⁹

X-ray crystallographic structure determination

The three-dimensional structure of **8** in complex with its target HIV-1 PR was determined by the method of molecular replacement to a resolution of 2.5 Å and refined to a final *R*-factor of 18.5%. Crystals of **8** and HIV-1 Q7K protease were grown from 1.0 M NaCl, 20 mM NaOAc pH 5.4, 1.0 mM EDTA, and 1.0 mM DTT.⁹ Diffraction intensities were measured, using CuKα radiation from a Rigaku generator at 30 mA and 40 kV and a Siemen's X-1000 position-sensitive area detector, from a single crystal with orthorhombic plate morphology and dimensions of 100 μm × 50 μm × 20 μm. Data were reduced and scaled in space group P2₁2₁2₁, *a* = 53.07 Å, *b* = 60.95 Å, *c* = 63.24 Å, using the Siemen's software to give an overall *R*_{sym} of 8.1% for 14,215 observations of 6594 unique reflections between 32 and 2.5 Å with an overall completeness of 87%. Molecular replacement proceeded with a probe model of HIV-1 PR determined in our laboratory⁹ with the ligands and waters removed, using the program X-PLOR.²⁰ The initial *R*-factor for all data between 7.0 and 2.5 Å was 38.7%. Rigid body minimization reduced the *R*-factor to 34.3% and least-squares positional refinement gave an *R*-factor of 22.6%. At this point, a difference density map calculated using terms (Fo–Fc)_{calc} clearly defined the atomic positions of **8** and an associated water molecule in the active site.

The Fo–Fc electron density map was used to position the atomic model of **8** in the active site of HIV-1 PR. The fluorenyl ring system was centered in the electron density filling subsite S2 and the remainder positioned using the allowed dihedral rotations. Least-squares positional refinement, simulated annealing refinement,

hand rebuilding, and inclusion of 83 structural water molecules resulted in a final *R*-factor of 18.5% for all data between 7.0 and 2.5 Å.

Acknowledgements

The work reported here was supported by National Institutes of Health Grant GM39552. F.M. was the recipient of fellowship 70361-14-RF from the American Foundation for AIDS Research.

References

1. Popovic, M.; Sarngadharan, M. G.; Read, E.; Gallo, R. C. *Science* **1984**, *224*, 497.
2. Kramer, R. A.; Schaber, M. D.; Skalka, A. M.; Ganguly, K.; Wong-Staal, F.; Reddy, E. P. *Science* **1986**, *231*, 1580.
3. Darke, P. L.; Nutt, R. F.; Brady, S. F.; Garsky, V. M.; Ciccarone, T. M.; Leu, C. T.; Lumma, P. K.; Freidinger, R. M.; Veber, D. F.; Sigal, I. S. *Biochem. Biophys. Res. Commun.* **1988**, *156*, 297.
4. Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R. A. F.; Scolnick, E. M.; Sigal, I. S. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4686.
5. Boehme, R. E.; Borthwick, A. D.; Wyatt, P. G. *Ann. Reports Med. Chem.* **1995**, *30*, 139.
6. Wlodawer, A.; Erickson, J. W. *Ann. Rev. Biochem.* **1993**, *62*, 543.
7. DesJarlais, R. L.; Seibel, G. L.; Kuntz, I. D.; Ortiz de Montellano, P. R.; Furth, P. S.; Alvarez, J. C.; DeCamp, D. L.; Babé, L. M.; Craik, C. S. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6644.
8. De Voss, J. J.; Sui, Z.; DeCamp, D. L.; Salto, R.; Babé, L. M.; Craik, C. S.; Ortiz de Montellano, P. R. *J. Med. Chem.* **1994**, *37*, 665.
9. Rutenber, E.; Fauman, E. B.; Keenan, R. J.; Fong, S.; Furth, P. S.; Ortiz de Montellano, P. R.; Meng, E.; Kuntz, I. D.; DeCamp, D. L.; Salto, R. et al. *J. Biol. Chem.* **1993**, *268*, 15343.
10. Sui, Z.; Salto, R.; Li, J.; Craik, C. S.; Ortiz de Montellano, P. R. *Bioorg. Med. Chem.* **1993**, *1*, 415.
11. McPhee, F.; Caldera, P. S.; Bemis, G. W.; McDonagh, A. F.; Kuntz, I. W.; Craik, C. S. *Biochem. J.* **1996**, *320*, 681.
12. MDL Information Systems, Inc., San Leandro, CA, USA.
13. The crystal structure coordinates are available from the authors.
14. Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B.; Wlodawer, A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805.
15. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
16. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Prot. Eng.* **1995**, *8*, 127.
17. Yu, Z.; Caldera, P.; McPhee, F.; De Voss, J. J.; Jones, P. R.; Burlingame, A. L.; Kuntz, I. D.; Craik, C. S.; Ortiz de Montellano, P. R. *J. Am. Chem. Soc.* **1996**, *118*, 5846.
18. Pichuantes, S.; Babé, L. M.; Barr, P. J.; DeCamp, D. L.; Craik, C. S. *J. Biol. Chem.* **1990**, *265*, 13890.
19. Babé, L. M.; Craik, C. S. *Antimicrob. Agents Chemother.* **1994**, *38*, 2430.
20. Brunger, A. T. *J. Mol. Biol.* **1988**, *203*, 803.

(Received in U.S.A. 21 November 1996; accepted 11 February 1997)